34th Annual International Conference of the IEEE EMBS
San Diego, California USA, 28 August - 1 September, 2012

The role of vision processing in prosthetic vision

Nick Barnes, Xuming He, Chris McCarthy, Lachlan Horne, Junae Kim, Adele Scott and Paulette Lieby

Abstract— **Prosthetic vision provides vision which is reduced in resolution and dynamic range compared to normal human vision. This comes about both due to residual damage to the visual system from the condition that caused vision loss, and due to limitations of current technology. However, even with limitations, prosthetic vision may still be able to support functional performance which is sufficient for tasks which are key to restoring independent living and quality of life. Here vision processing can play a key role, ensuring that information which is critical to the performance of key tasks is available within the capability of the available prosthetic vision. In this paper, we frame vision processing for prosthetic vision, highlight some key areas which present problems in terms of quality of life, and present examples where vision processing can help achieve better outcomes.**

I. Introduction

In 1896, d'Arsonval demonstrated inducing "phosphenes and vertigo" using electrical stimulation, showing the possibility of prosthetic vision by electrical stimulation. Recently, results are emerging showing possible clinical benefit from clinical studies from two groups. Humayun et al, [12] reported a best subject result of 1.8 logMAR visual acuity along with motion discrimination and orientation and mobility results improved significantly over system off. Zrenner et. al [24] reported letter reading and a best visual acuity of 1.69 logMAR with a 1500 active microphotodiodes implant.

Recent important results (e.g., [12], [24]) show improved vision performance for individuals with little or no vision following the implantation of a prosthetic vision device. Such results represent a significant step in the emergence of implantable prosthetic vision - a prosthetic vision which is stable, and sufficient to show visual results.

The US definition of legal blindness is 20/200 visual acuity or a visual field of 20 degrees or less,[1] which prosthetic vision is some way from. A key barrier to full restoration is that normal vision has high acuity, can perceive over a large field of view, and the dynamic range over which contrasts can be discerned is large. However, individuals still may have effective functional abilities despite being legally blind. Prosthetic vision may also improve quality of life without restoring normal vision - through supporting key tasks such as orientation and mobility, face recognition and communication, and reading. The key role for vision processing in prosthetic vision is to deliver more effective performance of tasks that are important to quality of life, given restricted visual function. Specifically, given a restricted number of phosphenes, and a restricted dynamic range of levels on these phosphenes, vision processing must enable key visual information to be retained in reduced resolution to allow the performance of functional tasks. This can be achieved by extracting key information for particular tasks from incoming high resolution image streams, then ensuring this is preserved in the resulting reduced bandwidth prosthetic vision image.

This paper defines vision processing for prosthetic vision, and its role, particularly in supporting functional vision. Vision processing for prosthetic vision is complementary to developments in other aspects of prosthetic vision research. We also look at the gaps between current prosthetic vision and the needs of individuals with low vision, and how vision processing may help facilitate better outcomes.

II. Components of prosthetic vision

In this paper we are concerned with stimulating implantable visual prosthetic devices.[2] Most implantable visual prosthetic devices perform neural stimulation of the human visual system, and use some type of electronic photosensors to recover the luminance of the visual scene. This allows the possibility for the device to perform some processing of the incoming visual information. Most implantable stimulating visual prosthetic devices use electrical stimulation of the human vision system, beginning with cortical stimulation in the late 1960s [2]. There has also been optic nerve stimulation [21], [3], and several methods of retinal stimulation which are generally described by the anatomical position of the stimulator. This includes epi-retinal (e.g., [16], [17]), sub-retinal (e.g., [24] trans-scaleral (e.g., [9]), and supra-choroidal (e.g., [20]). The major exception to electrical stimulation is the optogenetic approach, where neurons are modified so that they become photosensitive [8].

All proposed electrical stimulation devices require some form of electronic photoreceptive device, mostly this is an external camera. However, [24] makes use of photodiodes directly mounted on the implantable stimulator. Both external camera, and eye resident photodiodes have been shown to be effective in human trials. Current proposals for visual prosthetics using optogenetics include an external camera and a projecting device outside the eye to concentrate light sufficiently for activation [7]. Most current prosthetic vision devices include an external camera and a wearable vision processor outside the body. Although it is more complex to integrate vision processing with electronic photoreceptors, it is possible, so vision processing can be incorporated when there is no external camera.

[1] http://www.eeoc.gov/facts/blindness.html

[2] We exclude passive devices, e.g., implantable miniature telescope [5].

978-1-4577-1787-1/12/$26.00 ©2012 IEEE

III. Vision processing for prosthetic vision definition

Vision processing takes signals from incoming electronic photosensors, makes some modifications to those signals, and transmits them to the stimulation device. We propose here, that in general, vision processing for prosthetic vision is a mapping from a high resolution, high dynamic range incoming image which can be captured at high frequency, to a stimulation device with lower resolution, lower dynamic range, and potentially lower frequency output. There may also be the possibility of other sensors, such as range, or GPS also being used to supplement this information. We define this relationship as:

$$\phi = f(\psi), \tag{1}$$

f is a mapping from an input image stream, to a set of phosphenes, both of which can be described as consisting of a set of visual fields with two spatial and one temporal dimension, $\psi = \psi_{xt} = (I_x, I_t), \phi = \phi_{xt} = (P_x, P_t)$.

The output image set is spatially discrete, so we may define: P_x as a set of n_ϕ phosphenes in two spatial dimensions, $P_x = \{0, 1, ...n_\phi\}$ (note this is unlikely to be a regular grid in perceptual space, e.g., see [13]) and, $P_t \in \{0, 1, ...p_\phi\}$. The stimulation cycle will be finite, and may not be synchronous, but let us define this as a discrete output. Also, for each individual phosphene $\phi_i \in \phi_{xt}$, the value of ϕ_i is over a restricted range, $\phi_i \in \{0, 1, ..l_\phi\}$, where l_ϕ is the maximum number of discriminable levels of dynamic range of a phosphene. Note this may vary per phosphene.

By calibration, we may define any incoming camera rig configuration as a finite set of discrete input image visual fields (pixels) in two spatial dimensions: thus, I_x is a set of n_ψ image visual fields in two spatial dimensions, $I_x = \{0, 1, ...n_\psi\}$, which may be in a linear grid in input space. If we consider some finite interval of time, then the visual fields will be sampled a finite number of times. Let us assume that each visual field may be treated as taking the same number of samples during this time, so $I_t \in \{0, 1, ...p_\psi\}$. Further, each ψ_{xt} is discrete over some finite dynamic range of imaged luminance, therefore we may define $\psi_{xt} \in \{0, 1, ..l_\psi\}$, where l is the number of levels of dynamic range. Assume that an appropriate input device has been chosen so that ϕ is not oversampled in any dimension. Then we have $n_\phi \leq n_\psi$, $p_\phi \leq p_\psi$, and $l_\phi \leq l_\psi$.

IV. Current limitations on technologies

The number of phosphenes that can be separately induced by current generation implantable visual prosthetic devices is low compared to the resolution in number of pixels of standard current generation mobile phone cameras. Indeed, although there are many photoreceptors in normal human vision, the number of retinal ganglion cells, the axons of which make up the optic nerve, is low relative to cameras. Within the limitations of the device, it is generally advantageous to have more than one measurement of scene luminance per phosphene as this is more robust to input noise. In general, for external cameras and current generation prosthetic vision devices, the number of spatial samples of incoming light will be substantially greater than the spatial resolution of the prosthetic vision induced image. That is $n_\phi < n_\psi$. In reports to date, the number of levels that implantees can reliably discriminate is at most around 10 [13], [22]. Most electronic visual sensors have significantly larger dynamic range than this, so $l_\phi < l_\psi$. In the case of the temporal dimension, modern cameras can sample the scene quickly, however, cameras that have a much greater speed than human flicker fusion tend to be limited to specialist applications, and so may be less power efficient. [22] presents data that suggests prediction of sequential events improved after delays of 100 to 200 ms. 5 Hz would be a down-sample from current camera image streams, however, perhaps better times can be achieved. Further, the difference is not so large from standard sensing devices compared to resolution and dynamic range, so it is less clear there will be significant down-sampling in the temporal domain.

Thus, we define prosthetic vision processing as a down-sampling spatially and in dynamic range as per Equation (1). The task of vision processing is to preserve information in this down-sampling operation that is important to the functional abilities of implantees.

V. Gaps to human vision

Human vision has peak foveal sampling of incoming light of around 120 cycles per degree [23], and has a large field of view. Sampling is not of uniform high acuity, but, outside the foveal area reduces logarithmically in resolution [18]. Human vision incorporates fixatiion, which directs the fovea to areas of interest to allow effective high resolution over a substantial field of view.

Human vision also effectively has a large dynamic range over which it can perceive light. By dark and light adaptation of the retina [1] the eye is able to operate effectively in bright sunshine into quite dark conditions. Full adaptation between these conditions can take significant time. In a single fraction of a second, the dynamic range of light perception is greatly reduced, however, it is still significant. In comparison, the approximately 10 levels of distinguishable brightness demonstrated in current trials of implantable visual prosthetic devices is greatly limited.

Normal human vision can use binocular disparity to infer the distance of close objects, and can infer the range of distant objects through motion over time. The ability to control locomotion by cues from optical flow, such as centering and landing, without requiring absolute recovery of depth information have been well demonstrated [19].

However, even when bilateral implants are available, it is uncertain whether electrical stimulation will be able to restore ocular dominance columns which are associated with depth perception [4]. The performance of depth perception is closely correlated with visual acuity. For optical flow, computational models such as spatio-temporal derivatives and feature tracking critically depend on the precision of encoding of spatial contrast and so are sensitive to input dynamic range and spatial density. Thus, we can expect that the ability of prosthetic vision to infer depth is impaired relative to the abilities of normal human vision.

VI. Results for vision processing in prosthetic vision

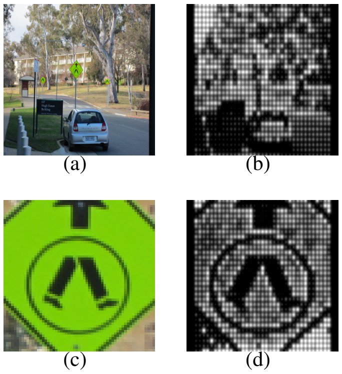
We now give some illustrative examples of our work demonstrating that improved results can be shown by prosthetic vision (using a simulation of prosthetic vision) above what would be expected from the direct functional abilities of the corresponding prosthetic vision.

A. *Visual fixation, face recognition, sign reading*

Human vision uses visual fixation to most effectively utilize the high resolution of the fovea. In prosthetic vision, given a spatial downs-ample from the input image, one may digitally zoom (perhaps also optically) to items of interest such as faces and signs. This allows the whole prosthetic vision resolution to be devoted to recognition, allowing vision processing to perform fixation and tracking.

Figure 1 shows that such an effect can help facilitate recognition of signs, as the result of using a sliding window-based object detector, using HOG feature descriptors (Dalal and Triggs [6]). This was trained using a large dataset of around 20,000 images of Australian pedestrian crossing sign images and other outdoor scenes as negative examples.[3] Figure 1 shows the raw image, and its phosphenized version; the second row shows the detected window only zoomed to full size along with its corresponding phosphene image. More detail can be seen in [11].

Human fixation plays a key role in face and facial expression recognition, particularly in dynamic scenes. The effect of face zooming is shown in Figure 2. In the system implemented, the face is zoomed once selected and tracked until the user disengages the interaction. This result was reported in [10].


(a) (b)

(c) (d)

Fig. 1. Automated detection and zooming of signs in prosthetic vision.

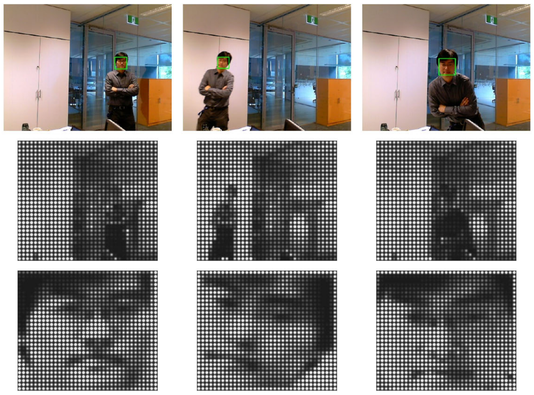


Fig. 2. Comparison of simulated prosthetic vision representation without face fixation versus with face fixation. Top row: The original high resolution input image frame at varying distances; Middle row: the phosphenized images with 35x30 simulated phosphenes; Bottom row: the detected and zoomed face region.

B. *Orientation and Mobility*

Given limited dynamic range, the ability to perceive small trip hazards can be impaired if they are of low contrast, particularly in the reduced dynamic range of prosthetic vision. Vision processing may augment the representation to ensure obstacle visibility from its background despite the differences in intensity (or depth if depth is represented on phosphenes instead) being insufficient to appear under expected quantization.

In [15] we demonstrated a system for finding the ground plane and ensuring that ground-based obstacles are apparent in the visual scene. Here, the ground plane was detected in disparity images taken by a stereo rig mounted on a skate board helmet and worn by the participant. The approach took particular care to find boundaries of objects with the ground plane, including the walls and trip hazards. The scene can be represented as a depth image to overcome problems of depth perception in low dynamic range visualizations, and the contrast of these boundaries was increased, so that potential trip-hazard obstacles pop-out of the visual scene. Figure 3 shows how a potential trip-hazard obstacle of low contrast can be difficult to see in a regular simulation of prosthetic vision, but that using an augmented depth representation it can be clearly differentiated.

It is necessary to evaluate the performance of vision processing algorithms for prosthetic vision before deployment. One way to do this is perform this evaluation using simulation software with normally sighted participants. For this purpose, we have developed real-time software that allows for customization of input image streams and rendered phosphene streams [14]. This simulated prosthetic vision software was used to produce the simulated prosthetic vision images shown in this paper.

VII. Conclusion

For any level of visual function that can be provided by prosthetic vision, the role of vision processing for prosthetic

[3] Thanks to Gary Overett of the Automap project for his assistance with sign detection.

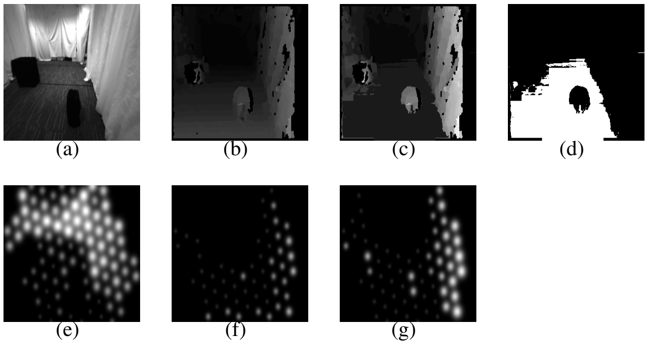


Fig. 3. The augmented ground-plane and alternative representations of a visual scene. (a) corridor intensity image (b) disparity image (c) the disparity image with an augmented ground plane and boundaries (d) the ground plane mask as recovered. (e), (f) and (g) show phosphene images corresponding to the intensity, depth, and augmented depth images respectively with 98 phosphenes and low dynamic range. Note the visibility of the trip hazard obstacle in (g).

vision, is to deliver functional ability that is improved by ensuring key information for tasks is preserved. We defined prosthetic vision as a mapping from input visual information to a stimulated pattern in the human visual system, which is down-sampled spatially and in dynamic range. Computer vision techniques may be used to select key information from incoming image streams, and ensure this is preserved. Examples of this were given with highlighting obstacles in limited dynamic range and using detection and tracking techniques to replace human fixation by zooming to key objects in the scene, and keeping them steady. This demonstrates a key role for vision processing in prosthetic vision.

References

[1] H Aubert. *Physiologie der Netzhaut*. Morgenstern, 1864.

[2] G S Brindley and W S Lewin. The sensations produced by electrical stimulation of the visual cortex. *Journal of Physiology*, 196(2):479–493, May 1968.

[3] X Chai, L Li, K Wu, C Zhou, P Cao, and Q Ren. C-sight visual prostheses for the blind. *IEEE Engineering in Medicine and Biology Magazine*, 27(5):20–28, Sept-Oct 2008.

[4] D B Chklovskii. Binocular disparity can explain the orientation of ocular dominance stripes in primate primary visual area (V1). *Vision Research*, 40(13):1765–1773, 2000.

[5] K A Colby, D F Chang, R D Stulting, and S S Lane. Surgical placement of an optical prosthetic device for end-stage macular degeneration: The implantable miniature telescope. *Archives of Ophthalmology*, 125(8):1118–1121, 2007.

[6] N Dalal and B Triggs. Histograms of oriented gradients for human detection. In *CVPR '05 Int Conf on Computer Vision and Pattern Recognition*. IEEE, 2005.

[7] P A Degenaar. Optogenetic visual prosthesis - engineering the optoelectronic stimulator. In *ARVO*, May 2011.

[8] K Deisseroth. Optogenetics. *Nature Methods*, 8:26–29, 2011.

[9] T Fujikado, M Kamei, H Sakaguchi, H Kanda, T Morimoto, Y Ikuno, K Nishida, H Kishima, T Mauro, K Konoma, M Ozawa, and K Nishida. Testing of semichronically implanted retinal prosthesis by suprachoroidal-transretinal stimulation in patients with retinitis pigmentosa. *Investigative Ophthalmology and Visual Science*, 52:4726–4733, Jun. 2011.

[10] Xuming He, Nick Barnes, and Chunhua Shen. Face detection and tracking in video to facilitate face recogntion with a visual prosthesis. In *ARVO*, May 2011.

[11] L Horne, N Barnes, X He, and C McCarthy. Object detection for bionic vision. In *2nd Int. Conf. on Medical Bionics: Neural interfaces for damaged nerves*, 2011.

[12] M S Humayun, J D Dorn, L Da Cruz, G Dagnelie, J-A Sachel, P E Stranga, A V Cideciyan, J L Duncan, D eliot, E Filley, A C Ho, A Satnos as A B Safran, A Arditi, L V Del Priore, and R J Greenberg for the Argus II Study Group. Interim results from the international trial of second sight's visual prosthesis. *Investigative Ophthalmology and Visual Science*, 2012. in press.

[13] M S Humayun, J Weiland, G Y Fujii, R Greenberg, R Williamson, J Little, B Mech, V Cimmarusti, G Van Boemel, G Dagnelie, and E deJuan Jr. Visual perception in a blind subject with a chronic microelectronic retinal prosthesis. *Vision Research*, 43:2573–2585, 2003.

[14] P Lieby, N Barnes, C McCarthy, N Liu, H Dennnett, J G Walker, V Botea, and A Scott. Substituting depth for intensity and real-time phosphene rendering: Visual navigation under low vision conditions. In *IEEE Int. Conf. of Engineering in Medicine and Biology Society (EMBC)*, Aug. 2011.

[15] C McCarthy, N Barnes, and P Lieby. Ground surface segmentation for navigation with a low resolution visual prosthesis. In *IEEE Int. Conf. of Engineering in Medicine and Biology Society (EMBC)*, Aug. 2011.

[16] D Nanduri, M Humayun, R Greenberg, M J McMahon, and J Weiland. Retinal prosthesis shape analysis. In *IEEE EMBC*, pages 1785–8, Aug. 2008.

[17] JF Rizzo, J Wyatt, J Loewenstein, S Kelly, and D Shire. Perceptual efficacy of electrical stimulation of human retina with a microelectrode array during short-term surgical trials. *Investigative Ophthalmology and Visual Science*, 44(12):5995–6003, 2003.

[18] E L Schwartz. A quantitive model of the functional architecture of human striate cortex with application to visual illustration and cortical texture analysis. *Biological Cybernetics*, 37:63–76, 1980.

[19] M V Srinivasan, S W Zhang, J S Chahl, E Barth, and S Venkatesh. How honeybees make grazing landings on flat surfaces. *Biological Cybernetics*, 83:171–83, 2000.

[20] G J Suaning, S Kisban, S C Chen, P J Byrnes-Preston, C Dodds, D Tsai, P Matteucci, S Herwik, J W Morely, N H Lovell, O Paul, T Stieglitz, and P Ruther. Discrete cortical responses from multi-site supra-choroidal electrical stimulation in the feline retina. In *IEEE EMBC*, pages 5879–5882, Aug. 2010.

[21] C Veraart, MC Wanet-Defalque, B Gerard, A Vanlierde, and J Delbeke. Pattern recognition with the optic nerve visual prosthesis. *Artificial Organs*, 27(11):996–1004, Nov. 2003.

[22] R Wilke, V-P Gabel, H Sachs, K-U Bartz-Scmidt, F Gekeler, D Besch, P Szurman, A Stett, B Wilhelm, T Peters, A Harscher, U Greppmaier, S Kibbel, H Benav, A Bruckmann, K Stingl, A Kusnyerik, and E Zrenner. Spatial resolution and perception of patterns mediated by a subretinal 16-electrode array in patients blineded by hereditary retinal distrophyies. *Investigative Ophthalmology and Visual Science*, 52(8):5995–6003, June 2011.

[23] D R Williams and H Hofer. *The Visual Neurosciences: Volume 1*, chapter Formation and Acquisition of the Retinal Image, pages 795–810. MIT Press, Cambridge Massachusetts, 2004.

[24] E Zrenner, KU Bartz-Schmidt, H Benav, D Besch, A Bruckmann, V-P Gabel, F Gekeler, U Greppaier, A Harscher, S Kibbel, J Kock, A Kusnyerik, T Peters, K Stingl, A Stett, P Szurman, B Wilhelm, and R Wilke. Subretinal electronic chips allow blind patients to read letters and combine them to words. *Proc. Royal Society of London B: Biological Sciences*, 278:1489–1497, 2011.